

文章编号: 1001-0920(2007)04-0418-05

视频序列中面向人的多目标跟踪算法

常发亮, 马 丽, 乔谊正

(山东大学 控制科学与工程学院, 济南 250061)

摘 要: 针对视频序列中人的跟踪问题, 提出一种基于运动检测的多目标跟踪算法. 跟踪系统由运动目标检测、关联矩阵建立、特殊情况判断及处理以及轨迹关联 4 部分构成. 提出一种基于改进的 c -均值聚类的自适应运动分割方法; 不同情况下建立不同的关联矩阵, 以准确判断实际场景状况; 对遮挡问题作出处理, 在两个目标遮挡不严重的情况下, 分别采用均值漂移算法对其进行跟踪. 实验结果表明, 该算法具有较强的鲁棒性, 能有效实现复杂场景下多目标跟踪.

关键词: 多目标跟踪; 运动检测; 轨迹关联; 遮挡; 均值漂移

中图分类号: TP391 **文献标识码:** A

Human oriented multi-target tracking algorithm in video sequence

CHANG Fa-liang, MA Li, QIAO Yi-zheng

(School of Control Science and Engineering, Shandong University, Ji 'nan 250061, China. Correspondent: Chang Fa-liang, E-mail: flchang@sdu.edu.cn)

Abstract: Aiming at tracking human in video sequence, a multi-target tracking algorithm based on moving detection is proposed. Tracking system is composed of four parts: Moving targets detection, association matrix forming, special cases estimation and handling, and track association. An adaptive moving segmentation method based on improved c -mean clutter is presented. Different association matrix in different condition is used to estimate the actual scene states accurately. Handling for occlusion is achieved, and in the case of mild occlusion of two targets, mean shift algorithm is applied to track them separately. Experimental results show that it is robust and has good multi-target tracking performances under complex background.

Key words: Multi-target tracking; Moving detection; Track association; Occlusion; Mean shift

1 引 言

面向人的视频跟踪技术是计算机视觉研究的难点之一, 在安全监控、虚拟现实、人机交互等领域有着非常重要的实用价值和广阔的发展前景.

著名的基于运动检测的面向人的跟踪系统有 Pfinder^[1]和 W^4 ^[2]. 其中 Pfinder 能够对室内的单个人体进行检测、跟踪和行为解释; W^4 则能在户外跟踪多个人体和人体各主要部分, 并可回答 Who, Where, When, What 等问题, 解释目标行为. 文献 [3-5] 是基于运动检测的可以解决遮挡问题的多目标跟踪算法, 利用目标的形状特征^[3]或颜色特征^[4]在遮挡期间跟踪目标, 或在遮挡后辨识原目标^[5].

本文针对人的跟踪中常出现的变形、遮挡以及多目标跟踪中常出现的目标合并、分裂、消失、新增等情况, 提出基于运动检测的多目标跟踪算法. 系统

主要由运动目标检测、关联矩阵建立、特殊情况判断及处理、轨迹关联等 4 部分构成. 首先利用背景差分进行运动目标检测, 提出基于改进的 c -均值聚类的自适应运动分割方法, 结合背景的实时更新来准确提取运动目标; 然后在不同情况下建立不同的关联矩阵, 并根据关联矩阵判断特殊情况的发生, 特别是对人的遮挡(对应合并、分裂)问题作出处理; 最后进行当前观测目标与已有轨迹的正确关联. 图 1 为跟踪系统的结构框图.

2 运动目标检测

背景差分法是运动目标检测的重要方法, 常见的背景模型有高斯模型^[1]、混合高斯模型^[6]及非参数模型^[7]等. 文献 [8] 利用双背景进行运动目标检测. 本文将改进的 c -均值聚类算法用于运动目标的自动分割, 结合背景实时更新保证目标检测的实时

收稿日期: 2005-11-16; 修回日期: 2006-01-03.

基金项目: 国家自然科学基金项目(60104009); 山东省自然科学基金项目(Z2005 G03).

作者简介: 常发亮(1965—), 男, 山东寿光人, 教授, 从事模式识别、机器视觉等研究; 马丽(1982), 女, 山东济宁人, 硕士, 从事计算机视觉、模式识别等研究.

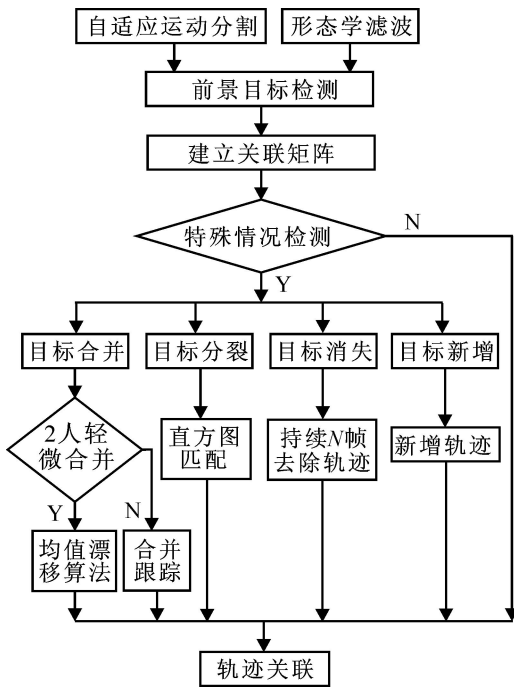


图 1 多目标跟踪系统结构框图

性和准确性,有效克服了通常的固定分割阈值难以适应场景变化的缺点.

2.1 自适应运动分割

在背景与当前帧的差分图像中,背景点对应值较小,运动目标点对应值较大.将差分值聚成两类,分别代表背景和目标,聚类的收敛相当于运动目标的分割.当前图像 I 与背景图像 M 的差分图像 D 为

$$D(x, y) = | I(x, y) - M(x, y) |. \quad (1)$$

选择背景的初始类心 $Z_1^{(0)}$ 为 $(5, 5, 5)$,前景目标的初始类心 $Z_2^{(0)}$ 为 $(50, 50, 50)$,前景与背景的背景初始标准差比为 $p^{(0)} = (3, 3, 3)$,采用基于马氏距离的改进的 c -均值算法对 RGB 三个颜色分量分别聚类.若点 (x, y) 的 RGB 三个颜色分量都属于背景,则该点被判断为背景点;否则为目标点.对于第 i 次迭代,

$$\begin{aligned} & \text{if } d(D_j(x, y), Z_{1j}^{(i)}) \times p_j^{(i)} < d(D_j(x, y), Z_{2j}^{(i)}), \\ & \quad j = R, G, B, \\ & \text{then } (x, y) = B, \\ & \text{elst } (x, y) = O, \end{aligned} \quad (2)$$

$$d(D_j(x, y), Z_{uj}^{(i)}) = | D_j(x, y) - Z_{uj}^{(i)} |, \quad j = R, G, B, u = 1, 2. \quad (3)$$

其中 $D_j(x, y)$ 为 (x, y) 点 j 颜色分量的差分值, $Z_{1j}^{(i)}$ 和 $Z_{2j}^{(i)}$ 分别为第 i 次迭代中背景类和目标类的第 j 个颜色分量的类心, B 代表背景, O 代表运动目标.为保证分割的有效性,加入约束条件使迭代在分割最优时停止,随着分界线向背景类移动,前景点的数量增加,在相邻两次增加数量最少的时候停止迭代,即

$$k^* = \arg \min_k (n(k) - n(k - 1)), \quad \text{if } n(k) > n(k - 1), \quad (4)$$

其中 $n(k)$ 为第 k 次迭代时的目标点数量.第 k^* 次的迭代结果作为运动目标分割的初始结果,然后进行去噪,将面积小于一定阈值的连通成分去除(实验中阈值选为 40),最后进行形态学滤波(用 3×3 结构元进行闭运算),填补初始分割中目标内部的洞,得到最终的分割结果.图 2 为运动目标的分割,其中,图 2(a) 为第 128 帧 R 分量的差分图像直方图(为清楚显示运动类,只取运动目标集中区域的直方图),图 2(b) 为第 128 帧图像,图 2(c) 为自适应运动分割结果,图 2(d) 为经过消噪和闭运算后的最终分割结果.

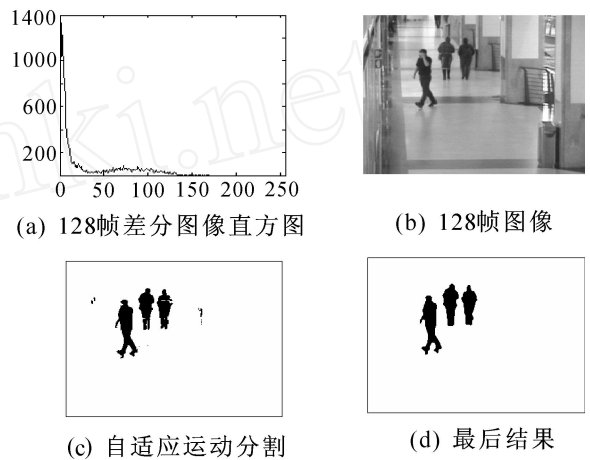


图 2 运动目标分割

2.2 背景实时更新

背景更新的基本思想是使当前帧有一定置信度的背景点参与更新.置信度通过两方面来估计,一是根据当前差分图像,由运动检测得到的非运动点不一定是背景点,如较弱的阴影、噪声点等,所以,只有差分值小于一定较小阈值并位于运动目标一定邻域外的点才具有第 1 方面较高置信度.将运动目标区域 O 用 5×5 结构元进行膨胀,记为 \hat{O} ,若

$$(x, y) \notin \hat{O} \text{ 且 } d(I(x, y), M(x, y)) < Th, \quad (5)$$

记 $f_1(x, y) = 1$,表明其差分值具有高置信度,其中 Th 为一个较小的阈值,实验中取为 6.

置信度的第 2 个方面是根据各点保持为背景点的历史信息,用矩阵 f_2 表示,只有持续 w 帧都为背景的点才可能参与背景更新^[5].初始帧 f_2^1 置 0,后续帧 f_2 的更新方法为

$$f_2^t(x, y) = \begin{cases} w, & (x, y) \in O; \\ w - 1, & (x, y) \in B \text{ 且 } f_2^{t-1}(x, y) = 0. \end{cases} \quad (6)$$

f_2^t 为 0 的点具有第 2 方面的高置信度,使同时

满足两个置信度要求的点参与背景更新,即

$$\begin{aligned} & \text{if } f1'(x, y) = 1 \text{ 且 } f2'(x, y) = 0, \\ & \text{then } M'(x, y) = \\ & \quad \alpha I'(x, y) + (1 - \alpha) \times M^{t-1}(x, y), \\ & \text{otherwise } M'(x, y) = M^{t-1}(x, y). \end{aligned} \quad (7)$$

其中 α 为遗忘因子,为防止过更新,实验中取 0.1.

3 关联矩阵建立及场景情况判断

3.1 关联矩阵的建立

关联矩阵是当前观测和已有轨迹进行关联的依据,是判断场景中各种特殊情况发生的依据.常用的关联矩阵可以基于连续帧目标质心间的距离^[5]或目标相交区域的面积^[3],本文将这两种性能互补的关联矩阵相结合,不同情况下采用不同的关联方法,以保证多目标跟踪的准确性.假设目标合并、分裂、消失、新增这 4 种情况在同一时刻不同时发生,则在连续帧间目标数量不变的情况下采用基于质心距离的关联矩阵;而当目标数量发生变化时,对应目标合并、分裂、消失或新增的发生,采用基于相交区域面积的关联矩阵.

基于质心距离的关联矩阵记为 Dis:以各轨迹和目标的距离为元素,行为已有轨迹 $\{T_1, T_2, \dots, T_n\}$, n 为已有的轨迹数量,列为当前帧各观测目标 $\{O_1, O_2, \dots, O_m\}$, m 为目标数量.基于 Dis 矩阵,建立判断矩阵 R ,对每一已有轨迹 T_i ,使其距离最近的目标 O_{j^*} 作为该轨迹的延续,置 R 中相应值为 1.

$$\begin{aligned} \text{Dis}(i, j) &= d(T_i, O_j) = \\ & \sqrt{(T_{ix} - O_{jx})^2 + (T_{iy} - O_{jy})^2}, \quad (8) \\ R(i, j^*) &= 1, \\ \text{if } j^* &= \arg \min_j (\text{Dis}(i, j)), \\ & i = 1, \dots, n. \end{aligned} \quad (9)$$

基于相交区域的关联矩阵记为 Area,以连续帧间各分割目标的相交面积为元素.若上帧目标 O_i^{t-1} (对应已有轨迹 T_i) 与当前观测目标 O_j^t 的相交面积大于一定阈值 T_2 (实验中取为 30 即符合实际),表明当前目标 O_j^t 为轨迹 T_i 的延续,置判断矩阵 R 中相应值为 1.

$$\text{Area}(i, j) = S(O_i^{t-1} \cap O_j^t), \quad (10)$$

$$R(i, j) = 1, \text{ if } \text{Area}(i, j) > T_2. \quad (11)$$

3.2 特殊情况判断及轨迹关联

若目标数量发生变化,表明发生了合并、分裂等特殊情况,具体由 Area 阵对应的 R 阵进行判断,方法如下:

Step1: 对 R 阵每一列 O_j , 计算该列的和

$$c_j = \sum_{i=1}^n R(i, j). \quad (12)$$

1) 若 $c_j = 0$, 表明 O_j 没有与其相关联的轨迹,为新增目标,令轨迹数量 $n = n + 1$, 增加 O_j 对应的新轨迹 T_n ;

2) 若 $c_j = 1$, 且对应 $R(i^*, j) = 1$, 则 O_j 与轨迹 T_{i^*} 关联;

3) 若 $c_j > 1$, 表明有 c_j 条轨迹合并为当前目标 O_j , c_j 条轨迹全部与 O_j 进行关联.

Step2: 对 R 阵每一行 T_i , 计算该行的和

$$r_i = \sum_{j=1}^m R(i, j). \quad (13)$$

1) 若 $r_i = 0$, 表明轨迹 T_i 中断, 其对应目标暂时消失, 因为目标可能被静止物体暂时遮挡, 所以只有该轨迹持续 N 帧都没有相关联的目标, 再结束该轨迹;

2) 若 $r_i = 1$, 并对应 $R(i, j^*) = 1$, 则 T_i 与目标 O_{j^*} 关联;

3) 若 $r_i > 1$, 表明轨迹 T_i 与 r_i 个目标关联, 分裂发生, 对每一个分裂出去的目标, 用直方图匹配方法将该目标与其合并前的轨迹进行关联.

4 遮挡情况下的多目标跟踪

遮挡是人的跟踪中常出现的难点问题, 多目标之间的遮挡对应于上述的合并和分裂情况. 将遮挡分为两种情况进行处理, 在两人轻微遮挡造成的合并情况下, 为保证跟踪的准确性, 采用基于颜色直方图的均值漂移算法分别跟踪两个目标, 其余情况下将所合并目标一起跟踪, 实时更新所有合并的轨迹, 并在遮挡过程中使合并轨迹保持同样的关联. 遮挡结束后再分别跟踪, 并与遮挡前的轨迹进行关联.

4.1 不严重遮挡情况下基于均值漂移算法的跟踪

对应于上述 Step1 中的情况 3), 若 $c_j = 2$, 表明有两条轨迹在当前时刻 t 合并, 设 T_a, T_b 两轨迹合并为目标 O_k , 若各目标的面积满足

$$S(O_k^t) > 0.8 \times (S(O_a^{t-1}) + S(O_b^{t-1})), \quad (14)$$

则判断遮挡不严重, 标记 $ms^t(a, b) = 1$, 用均值漂移算法^[9,10] 分别跟踪合并在一起的两个目标.

实时保存各目标的 HSI 三维颜色直方图 $q_u^t(i)$, $i = 1, \dots, m$, 其中 u 为三维颜色向量. 若 $ms^t(a, b) = 1$, 则分别以 O_a^{t-1} 和 O_b^{t-1} 的质心 $C_a(x_0, y_0)$ 和 $C_b(x_0, y_0)$ 为均值漂移算法的初始位置, 对跟踪窗内运动点标记 $f_i(x, y) = 1, i = a, b$, 非运动点标记 $f_i(x, y) = 0$, 按直方图 $q_u^{t-1}(i)$ 对跟踪窗内每一点 (x, y) 赋权值, 即

$$w_i^t(x, y) = q_u^{t-1}(i) \times (g(x, y) - u) \times f_i(x, y), \quad i = a, b. \quad (15)$$

其中: $g(x, y)$ 为 (x, y) 点在直方图中颜色索引值, $w_i^t(x, y)$ 为该点权值. 以 $C_i(x_0, y_0)$ 作为初始跟踪位

置,计算 t 帧各目标中心 $C_i(x_1, y_1)$ ($i = a, b$) 的均值漂移算法步骤如下^[10]:

Step1: 计算当前跟踪窗内各点 (x, y) 的权值 $w_i^t(x, y)$, $1 \leq x \leq m_i, 1 \leq y \leq n_i$. m_i 和 n_i 为各目标跟踪窗的长度和宽度.

Step2: 计算新的漂移位置 $C_i(x_1, y_1)$, 其中

$$x_1 = \frac{\sum_x x \times w_i^t(x, y)}{\sum_x w_i^t(x, y)},$$

$$y_1 = \frac{\sum_y y \times w_i^t(x, y)}{\sum_y w_i^t(x, y)}. \quad (16)$$

Step3: 若 $|C_i(x_1, y_1) - C_i(x_0, y_0)| < \epsilon$, 则停止计算; 否则将 $C_i(x_1, y_1)$ 作为 $C_i(x_0, y_0)$ 返回 Step1 迭代, 参数 ϵ 的设置应保证两次收敛到同一点.

4.2 遮挡结束后的轨迹关联

遮挡结束对应目标分裂, 关键是要判断分裂出的目标属于遮挡前的哪一条轨迹, 采用颜色直方图匹配的判断方法能适应目标易变形的特点. 设当前分裂出的目标为 O_j , 直方图为 p_u , 各合并目标 (设有 k 个) 直方图为 $\{q_u(i), i = 1, \dots, k\}$. 用 Bhattacharyya 系数的离散形式表示直方图的匹配度^[9]:

$$(i) = [p_u, q_u(i)] = \frac{1}{\sqrt{p_u \times q_u(i)}} \sum_{u=1}^{N_H \times N_S \times N_I} \sqrt{p_u \times q_u(i)},$$

$$i = 1, 2, \dots, k, \quad (17)$$

$$i^* = \arg \max_i (i). \quad (18)$$

选择 T_{i^*} 为该目标在遮挡前的轨迹. 若大的合并分裂为几个小的合并, 也是通过直方图匹配使每个目标与其所属的小的合并关联^[4].

5 实验结果及分析

5.1 遮挡不严重情况下对合并目标分别采用均值漂移算法跟踪

图 3 左右分别对应运动分割图 and 原图像. 其中: 图 3(a) 为遮挡前的跟踪; 图 3(b) 为遮挡开始时基于运动分割的跟踪, 两个目标发生合并而作为一个整体, 其质心与两个目标的质心差别较大, 跟踪不准确; 图 3(c) 为采用均值漂移算法对两个目标分别跟踪的结果, 可见跟踪效果更好; 图 3(d) 为目标遮挡结束 (分裂) 后基于运动分割的跟踪.

5.2 存在严重遮挡的多目标跟踪

图 4 为存在多目标严重遮挡的跟踪过程, 左右分别为运动分割图 and 原图像, 其中各曲线分别为各目标的运动轨迹. 图 4(a) 为其中 2 人发生轻微遮挡,

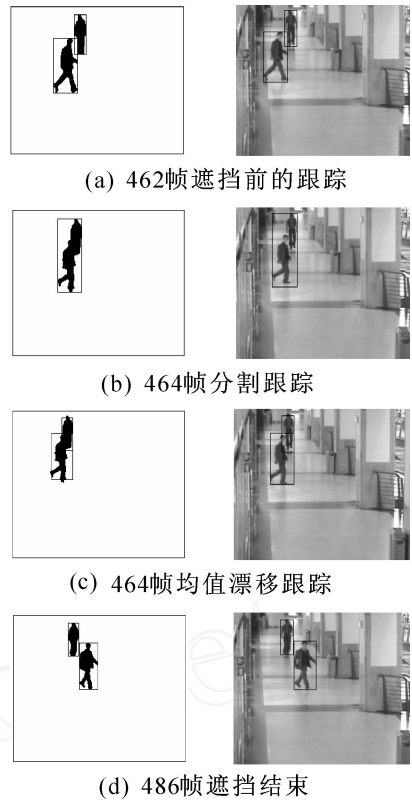


图 3 遮挡不严重情况下对合并目标分别采用均值漂移跟踪

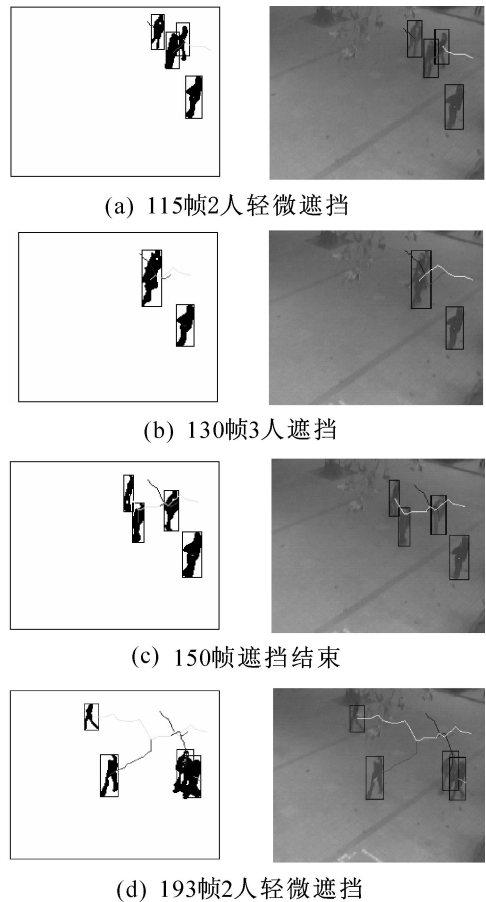


图 4 存在严重遮挡的多目标跟踪
对这两个目标的跟踪采用均值漂移算法, 其余目标仍按运动分割跟踪; 图 4(b) 为其中 3 人的严重遮

挡,将其作为一个整体进行跟踪,3人的轨迹暂时合并为一条轨迹;图4(c)为遮挡结束,3人完全分裂开,用直方图匹配算法分别辨识分裂目标,并与遮挡前的相应轨迹进行关联;图4(d)为2人又发生轻微遮挡,使用均值漂移算法对其进行跟踪。

5.3 复杂场景下的多目标跟踪

图5所示为复杂场景下的多目标跟踪结果,并实时描述各目标的运动轨迹。图5(a)为在61帧时出现新目标(最右侧目标),相应的为其新增轨迹;图5(b)为3个目标发生合并,一起进行跟踪;图5(c)在图像左上又出现新增的目标,并且黑色轨迹对应目标接近消失;图5(d)为已经确定黑色轨迹对应目标消失,去除其对应轨迹,并且左下方又出现新的目标;图5(e)为跟踪结束时的场景。可见,提出的关联矩阵和轨迹关联算法能对各种特殊情况进行准确判断并保证正确的跟踪。



(a) 61帧新增目标 (b) 108帧合并 (c) 146帧新增目标



(d) 166帧新增目标 (e) 320帧跟踪结束

图5 复杂场景下的多目标跟踪

6 结 论

针对多人跟踪中常出现的遮挡等难题,提出了基于运动目标检测的多目标跟踪算法。其特点有:1)基于改进 c -均值聚类的自适应运动目标检测方法,有效克服了固定分割阈值不能适应场景变化的缺点,并实时进行背景更新;2)不同情况下采用不同的关联矩阵,可保证场景情况判断的准确性;3)对于常出现的两人不严重遮挡问题,采用基于颜色直方

图的均值漂移算法对合并的目标分别跟踪,可保证跟踪的准确性;4)提出的轨迹关联方法和采用直方图匹配对分裂目标进行辨识,可保证多目标跟踪的可靠性。

参考文献(References)

- [1] Wren Christopher Richard, Azarbajehani Ali, Darrell Trevor, et al. Pfunder: Real-time tracking of the human body[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780-785.
- [2] Haritaoglu Ismail, Harwood David, Davis Larry S. W⁴: Real-time surveillance of people and their activities[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809-830.
- [3] Marecenaro L, Ferrari M, Marchesotti L. Multiple object tracking under heavy occlusions by using Kalman filter based on shape matching[C]. IEEE Int Conf on Image Processing. New York, 2002: 341-344.
- [4] Stephen J McKenna, Sumer Jabri, Zoran Duric, et al. Tracking groups of people[J]. Computer Vision and Image Understanding, 2000, 80(1): 42-56.
- [5] Tao Yang, Stan Z Li, Quan Pan, et al. Real-time multiple objects tracking with occlusion handling in dynamic scenes[C]. Computer Vision and Pattern Recognition. San Diego, 2005: 970-975.
- [6] Lee Dar-Shayang. Effective gaussian mixture learning for video background subtraction[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827-832.
- [7] Elgammal Ahmed, Harwood David, Davis Larry. Non-parametric model for background subtraction[C]. European Conf on Computer Vision. Dublin, 2000: 751-767.
- [8] Elias Herrero-Jaraba, Carlos Orrite-Urunuela, Jesus Senar. Detected motion classification with a double-background and a neighborhood-based difference[J]. Pattern Recognition Letters, 2003, 24(1): 2079-2092.
- [9] Comaniciu Dorin, Ramesh Visvanathan, Meer Peter. Kernel-based object tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
- [10] Bradski Gray R. Computer vision face tracking for use in a perceptual user interface[J]. Intel Technology J, 1998, 2: 1-15.

(上接第417页)

- [16] Li X, Liu L. Peer trust: Supporting reputation-based trust in peer-to-peer communities[J]. IEEE Trans on Knowledge and Data Engineering, 2004, 16(7): 843-857.
- [17] Breese J, Hecherman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]. Proc of the 14th Conf on Uncertainty in Artificial

Intelligence. San Francisco: Morgan Kaufmann Publishers, 1998: 43-52.

- [18] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]. Proc of the 10th Int World Wide Web Conf. Hong Kong: ACM Press, 2001: 285-295.